PROVIDING INFORMATION ON THE WORLD WIDE WEB

John H. Porter

Department of Environmental Sciences, Clark Hall, University of Virginia,
Charlottesville, VA 22903

Abstract. The capabilities provided by the World Wide Web (WWW) offer an opportunity for ecological researchers to share information resources. The hypertext markup language (HTML) is used to create documents for display on the WWW. HTML documents can be created using various tools from general text editors to more specialized programs. General principles of WWW page design that can be applied to improve content and usability include anticipating user needs and avoiding features that unnecessarily increase needs for network bandwidth. Search and indexing tools for WWW pages can be used to improve access to information. WWW pages can also be used to solicit information from users via on-line forms. Making HTML documents available on the WWW is accomplished by placing them on a server, which may be locally administered or available commercially.

GETTING STARTED - DEMYSTIFYING HTML

Information provided on the World Wide Web (WWW) comes in many multimedia forms. WWW software supports display of text, graphics, animation and sound. Despite its sophisticated and extensive capabilities, the underlying technologies are surprisingly simple. Web pages are written in the "HyperText Markup Language" (HTML). HTML files are simple ASCII text files that can be created by any software editor. They take the form of text enclosed in HTML "tags." The HTML "tags" are text enclosed in <>s. For example, <B> starts boldface and </B> turns off Boldface "a <B>test</B>" would appear as "a **test**."

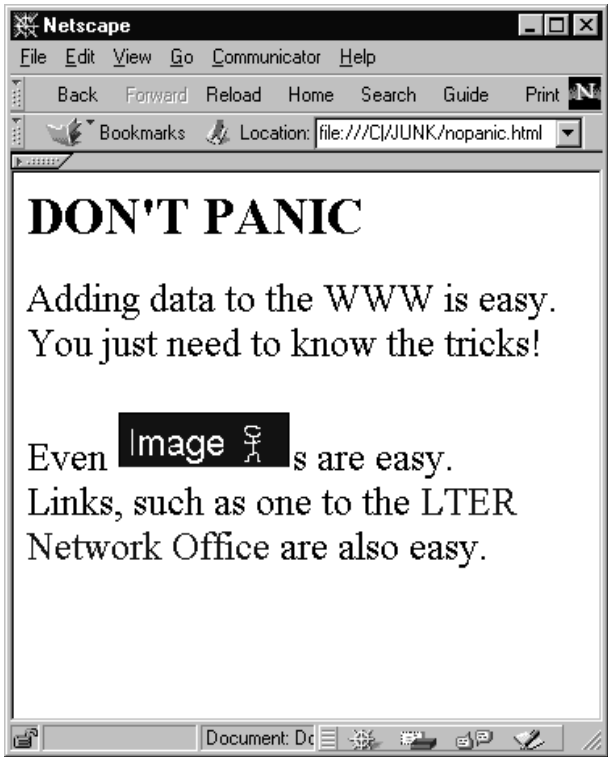
A more extensive example that includes both an image and a linkage to a WWW server is:

```
<H1>DON'T PANIC</H1>
<P>Adding data to the WWW is easy. You just need to know the
 tricks! </P>

<P>Even <IMG SRC=image.gif>s are easy. <BR> Links, such as one
 to the <A HREF="http://lternet.edu">LTER Network Office</A>
 are also easy.
</P>
```

Note that most tags occur as pairs. <P> and </P> are used to separate paragraphs. <H1> and </H1> delimit a heading, which is displayed in large type. There is an "anchor" <A HREF="http://lternet.edu"> </A> which outlines text ("LTER Network Office") that is displayed in blue. If clicked on in the browser, you would be transferred to the WWW site or document described in the HREF section of the anchor. Other tags occur as singletons.
 produces a line break. <IMG SRC="*name_of_graphics_file*"> inserts a graphics file. In the example the graphics file contains the word "image" in a blue box.

Figure 1. Illustration of page display (see text) from a WWW browser.



An excellent "HTML Table of Contents," assembled by Ian Graham, is available at: *http://www.utoronto.ca/webdoes/HTMLdoes/NewHTML/htmlindex.html*. However, it is not necessary for a creator of HTML documents to know or understand HTML because of the increasingly diverse range of software tools that create HTML documents. Recent versions of Netscape, for example, include a built-in editor for HTML documents. Similarly, Microsoft WORD allows you to "*Save As*.." an HTML document. WordPerfect for Windows and many other software packages include an HTML converter. There are also specialized packages, such as Microsoft FrontPage, that focus on WWW publishing.

If you see something interesting about the structure of a WWW page provided by someone else, it is possible to see exactly how he or she did it. WWW servers send copies of HTML files to the browser for interpretation and display. Using the "view source document" command you can look directly at the HTML source used to generate the page and see the commands responsible for creating the interesting page display and apply those same "tricks" to creating your own pages.

WWW PAGE DESIGN

Design of WWW pages is an art rather than a science. Nonetheless, there are some principles that can be applied to the creation of effective and informative WWW pages. The first is to know your audience! What types of information will they want? What types of information will they need most? The second is to use that knowledge to structure information on the server. Frequently requested information should be easy to locate. Information should be grouped into understandable categories to keep the number of menu entries at roughly seven or fewer so that the entire menu can appear on a single page.

Use of graphics is an area of both great opportunity and great hazard. Appropriate graphics can add interest and clarity to a page. However, they need to be used with caution. Graphics files (typically .GIF or .JPG files) are frequently large and WWW page displays can be significantly slowed by the inclusion of several graphics files. This problem is especially acute for users with slow network connections. Although they have improved dramatically in recent years, modem connections via telephone lines are tens to hundreds of times slower than direct connections to the Internet via a local area such as an ethernet. Thus a graphic image that may take only seconds to display at 10 megabits per second, can take several minutes to display at 56 kilobits per second.

Animated graphics place an additional load on the user's computer, which can significantly impact other applications. Graphical page backgrounds are an especially sensitive issue. They can be quite attractive on one system, but may display differently on other systems, which have different color capabilities. In some cases, the display of the background may be so poor as to make the overlaying page unusable. For this reason, it is a good idea to test any background on a variety of systems and software before employing it.

JAVA™ Applets are another opportunity. These applets are small programs that run on the user's system. They can add interactive characteristics to a WWW site that go beyond those available with HTML alone. However, like graphics, they can substantially increase the load imposed on the user's system and should therefore be used with caution.

USING A WWW SERVER

Adding pages to a WWW server is easy. It is merely a question of uploading the relevant HTML and graphics files to a directory on the server where they are accessible. Most educational institutions have servers that are available for use by faculty and students. On a university UNIX system, the process is often as simple as creating a directory named "public_html" in your home directory. Files placed in the "public_html" directory are then accessible over the WWW at the address: *http://address.of.server/~your_id/filename.html* where *address.of.server* is the network address of the computer (such as *poe.acc.virginia.edu*), *your_id* is the user-ID for your account and *filename.html* is the name of the specific file stored in your "public_html" directory. Commercial Internet Service Providers (ISP's) can also provide access to existing WWW servers.

You can also develop your own server. Full-featured server software is available for Windows 95™, Windows NT™, Unix and Macintosh systems, both as commercial software and as shareware. More limited servers are available for MSDOS. The only limit on a server computer is that it needs to have a stable network address and be accessible 24 hours a day (since users are seldom in a single time zone).

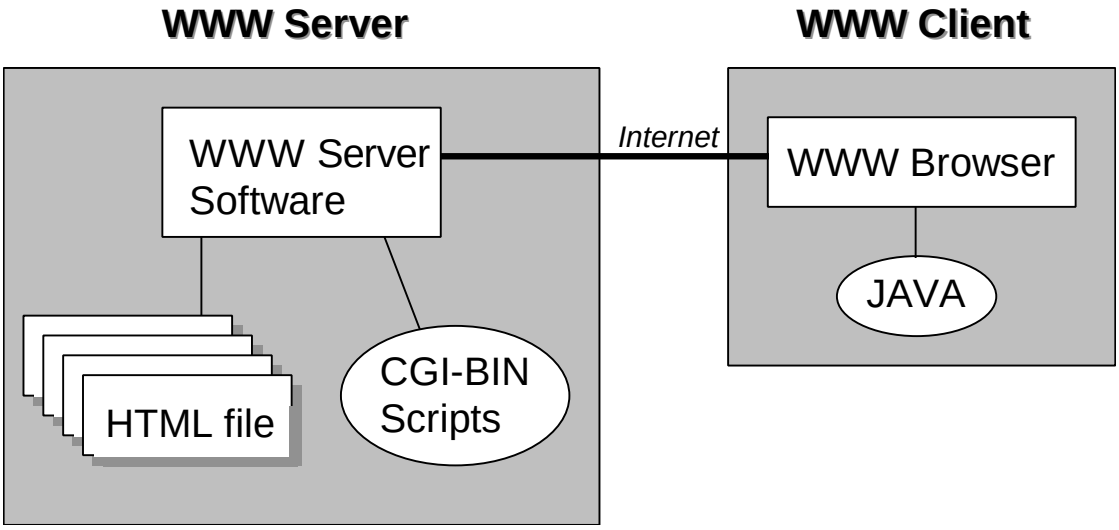
Server software for the WWW is relatively simple. A browser program on the client computer sends a request to the server for a particular file (such as an HTML document or graphics file). The server then sends a copy of the requested file back to the client for display. Indeed, it is one of the paradoxes of the WWW that the client browser program needs to be much more sophisticated than the server! Whereas a basic server simply needs to respond to requests for files (similar to File Transfer Protocol - FTP), a client browser (such as Netscape) needs to know how to convert the HTML files and graphics files into forms suitable for display on the screen. More sophisticated servers support Common Gateway Interface (CGI) scripts which dynamically generate WWW pages and support passwords and other security features. Additionally, they may include extensive logging capabilities to track use of individual pages.

In addition to the display of "static" HTML documents, the WWW can also display the output from programs. Some of these programs run on the server. Typically these programs are stored in the /cgi-bin directory on the WWW server. When invoked, they generate output that is then

returned across the network to the browser. These "cgi-bin" programs can be used to handle the input from WWW on-line forms and interface with database software.

Other programs run on the user's (client's) computer. Typically these are written in JAVA™ or ActiveX™ languages which incorporate special security features that limit what they are allowed to do on the client. This security is necessary because few of us would want to use the WWW if any link could download a program that would delete all our files! The relationship of the server and client (user) systems is displayed in Figure 2.

Figure 2. The relationship of the server and client (user) systems.



In deciding whether to use an existing server or create a new server, a critical issue is whether or not you need to use cgi-bin scripts. Most university and commercial servers restrict the use of server-side scripts because these place additional demands on the server computer and can create security holes.

SEARCHING AND FORMS ON THE WWW

The use of a WWW server is facilitated by tools that allow users to search for information without knowing where it is located on the server. There are several approaches that all involve using cgi-bin scripts (server-side programs) to access pre-computed indices. One approach uses full-text search engines. These create an index of all the words used in each WWW page. Searches yield links to all pages that contain a given search term. Examples of this type of search engine are WAIS (Wide Area Information Server), Glimpse/WebGlimpse™ and Excite™. Other types of search engines permit the user to search only certain fields (e.g., search for term only in title) such as Z39.50-compliant search engines and relational databases.

The utility of the WWW is greatly enhanced by using web pages (i.e., web forms) to collect information as well as to distribute information. HTML includes ways to create on-line forms. These forms feature a variety of different options for input, including text fields (both single line and text blocks), radio-buttons, check boxes and selection bars. The output from a form is an encoded string that includes the name of each field and the value that the user assigned to that field. Form output can be decoded by cgi-bin programs to create new data files or to interact with database software. At the Virginia Coast Reserve LTER site (*http://www.vcrlter.virginia.edu*) WWW forms are used to allow researchers to add entries to the site research calendar, research

abstracts, and annual reports. When coupled to a database, forms are used to update entries in the site personnel directory, biodiversity database and data catalogs. This greatly aids in developing scaleable solutions to common problems of site management by eliminating an input bottleneck. Any investigator with access to a WWW browser has the ability to update the databases.

SOURCES OF INFORMATION AND SOFTWARE

- HTML
  http://www.utoronto.ca/webdoes/HTMLdoes/NewHTML/htmlindex.html - HTML table of contents
- Server Software
  http://www.apache.org - Apache Server
  http://hoohoo.nesa.uiue.edu - NCSA WWW Server Software
  http://www.tucows.com - PC Networking Software
  http://www.microsoft.com - Microsoft WWW Server and Explorer Software
  http://www.netscape.com - Netscape WWW Server and Navigator Software
- Indexing and Searching
  http://www.goose.ycp.edu/~vkline/dragnet.html - Review on cataloging with lots of links
  http://www.glimpse.cs.arizona.edu/webglimpse/index.html - Web Glimpse Search Engine
  http://www.excite.com/navigate/ - Excite Web Site Search Engine